Antibodies That React with Predetermined Sites on Proteins

J. Gregor Sutcliffe, Thomas M. Shinnick
Nicola Green, Richard A. Lerner

Summary. Contrary to previous predictions, relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. Peptides capable of eliciting protein-reactive serums are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins nor to the amino or carboxyl terminals. As such, synthetic peptide immunogens are valuable for eliciting reagents with predetermined specificity that can be used for basic research. In addition, some synthetic peptides are capable of mimicking regions of virus proteins and eliciting immune responses in animals that are protective against the viral agents. Such peptides may thus serve as the basis for safe, chemically defined synthetic vaccines.

Progress in molecular biology and virology has always relied heavily on structural analyses of the components of biological systems. Because of advances in nucleic acid biochemistry during the last decade, culminating with the development of rapid DNA sequencing methods that allow the primary chemical sequences of genes to be read (*1*), we can now often approach a biological problem by identifying and sequencing the responsible genes rather than analyzing the relevant proteins. In fact, nucleotide sequence analysis of a new virus or gene is usually a most rewarding initial step, since it provides the information the virus or gene uses to accomplish its phenotype as well as a record of the evolutionary history of that nucleic acid. No fact is more important about a gene than its primary sequence; however, proteins are responsible for the execution of most biological processes.

Given the ease with which it was possible to generate primary sequence data for genes, what was needed was a way to link gene sequences to proteins, particularly when a protein was known only by the sequence of a gene. The obvious answer was to use an antibody, but the problem was how to make an antigen corresponding to a putative protein deduced from a string of nucleotides. One solution was to make a bacterium (or yeast) synthesize the novel protein and then raise antibodies to the bacterially synthesized protein. However, experience showed that although it was possible to make bacteria synthesize specific proteins if the genes for these proteins had been isolated, each new gene had to be handled individually and required many manipulations (*2*). If one were investigating a "gene" whose product was purely hypothetical, a laborious set of experiments might be necessary to achieve its bacterial expression.

An alternative to biological synthesis of antigens is chemical synthesis. However, most biologically interesting proteins whose sequences can be inferred from genes are in the 15,000 to 150,000 dalton range (15K to 150K), whereas chemical synthesis has practical limits in the 4K to 5K range except in exceptional circumstances. Furthermore, most studies of protein immunogenicity (including the work of Landsteiner, Crumpton, Benjamini *et al.*, Atassi and Suplin, Arnon and Sela and their colleagues, and Cebra and others) (*3*) indicated that small portions of a protein would, in general, be unlikely to elicit antiserums reactive against an intact protein [for a recent review, see (*4*)]. These studies predicted that immunogenic sites in small, intact proteins occurred about once every 5K to 10K and that these few loci relied on complex tertiary interactions between amino acid residues near each other in the protein tertiary structure, but distant in relation to the primary linear amino acid structure (so-called conformational determinants) (see Fig. 1). Therefore, it was generally believed that linear sequences in short peptides would not usually mimic these important sites. Recent experiments have challenged this prediction and have shown that small, chemically synthesized fragments of a protein can, in fact, elicit antibodies reactive with the native protein, thus allowing nucleic acid sequences to be parlayed quickly into biological experiments.

Antibodies to Synthetic Peptides React with Native Proteins

The belief that most antigenic determinants are conformational was first challenged in experiments with chemically synthesized protein fragments from the amino or carboxyl terminals of viral proteins whose sequences had been determined from nucleic acid studies. Antibodies to synthetic peptides [prepared as described in (*5*, *6*)] corresponding to the COOH-terminus of the envelope polyprotein of Moloney murine leukemia virus (MuLV) and the NH_2- and COOH-terminals of the simian virus 40 (SV40) transforming protein were found to be reactive with the native protein structures (*7*). That is, each reagent was able to precipitate the corresponding protein from extracts of virus-infected cells. In addition, the fact that antibodies to peptides are specific for predetermined sequences within the intact protein permitted analysis of the precursor of the MuLV envelope polyprotein that undergoes two stages of proteolytic cleavage necessary to generate mature viral proteins (*8*).

The first reports that protein-reactive antibodies could be elicited by synthetic peptides corresponding to fragments of proteins whose sequences were known only from nucleic acid studies (*7*) gave a clear indication that a powerful technology was now available (*9*). The possibility remained, however, that the utility of the technique would be confined to the terminals of proteins where carrier-coupled peptides might mimic their position in the intact molecule. We needed to know whether the technique would be applicable to all proteins and if it would be necessary to predict the naturally anti-

The authors are members of the Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037.

genic regions of proteins in order to select suitable synthetic peptides (*10*). We used the recently derived sequences of the hepatitis B virus surface antigen (HB_sAg) and the hemagglutinin gene of influenza virus type A (HA1) (*11*) to examine the general utility and expand our understanding of this technology. The two sequences offered different experimental situations. The HB_sAg was known to be a molecule whose immunogenicity is critically dependent on its native tertiary structure (*12*); furthermore, it is extremely hydrophobic. The crystallographic structure of influenza A hemagglutinin had recently been solved (*13*) and its dominant antigenic sites were well known (*14*). In addition, since both molecules are the primary targets of neutralizing antibody during natural infection they are of practical interest in dealing with human disease (*14–16*).

One question we needed to answer concerned the peptide characteristics required to elicit antibodies reactive with the native molecule: Are there restrictions in the chemical makeup or location within the protein structure? The hydrophobic HB_sAg glycoprotein consists of 226 amino acid residues. We synthesized 13 peptides using the HB_sAg amino acid sequence deduced from the hepatitis B genome sequence as a blueprint (*17*). These peptides were distributed throughout the primary sequence, but avoided in those regions that showed significant variability in the different viral isolates. Four were subsets of longer peptides. From the data obtained with these peptides we were able to formulate an elementary set of rules for selecting peptides capable of eliciting antiserums reactive with native proteins. Peptides that were extremely hydrophobic and those of six or fewer residues were ineffective; longer, soluble peptides, especially those containing proline residues, were effective. Antiserums against four of the six HB_sAg peptides in the latter category precipitated the HB_sAg protein in the viral Dane particles. Precipitation also occurred under conditions approximating physiological (saline solutions), indicating that antibodies to these peptides might be expected to bind antigen in vivo. This study showed that linear peptides from more than one region of a protein and, more important, not restricted to its NH_2- or COOH-terminus could elicit protein-reactive antibodies.

We also needed to know the relation between the sites represented by effective peptides and the antigenic determinants selected by the host in the course of a natural immune response against a virus or a protein. For influenza virus, those sites immunogenic during natural infection had been mapped by analysis of variants to four domains of the HA1 chain of the hemagglutinin molecule, whose three-dimensional structure was known from x-ray crystallographic studies (*13–15*). We studied 20 synthetic peptides, many of them overlapping, covering 75 percent of the HA1 primary sequence derived from the nucleotide sequence of a fragment of the influenza genome (*6*). In accordance with the rules formulated in the HB_sAg study (*17*), these peptides ranged in length from 8 to 39 residues, contained enough polar amino acids to render them soluble and, hence, easy to work with, and usually contained one or more proline residues. Some of the peptides fell within the known antigenic domains of HA1; others were clearly outside these domains. The peptides correspond to regions of the protein which in the crystal structure appeared as α-helices, β-sheets, and random coils. Antibodies to 18 of the 20 peptides reacted with HA1 (isolated by bromelain cleavage) or intact virus, demonstrating that sites in proteins accessible to peptide antibodies are more numerous than the few sites recognized in the course of a natural immune response. Furthermore, the information carried within a relatively short linear peptide is sufficient to elicit reactivity against a much larger protein molecule with a complex tertiary and quarternary structure. In more recent studies with peptides selected by the same rules outlined above, 12 of 12 peptides predicted from the MuLV polymerase gene (*18*) and 18 of 18 peptides from the rabies glycoprotein gene (*19*) elicited antibodies that precipitated their putative corresponding proteins. Therefore, it appears that, by

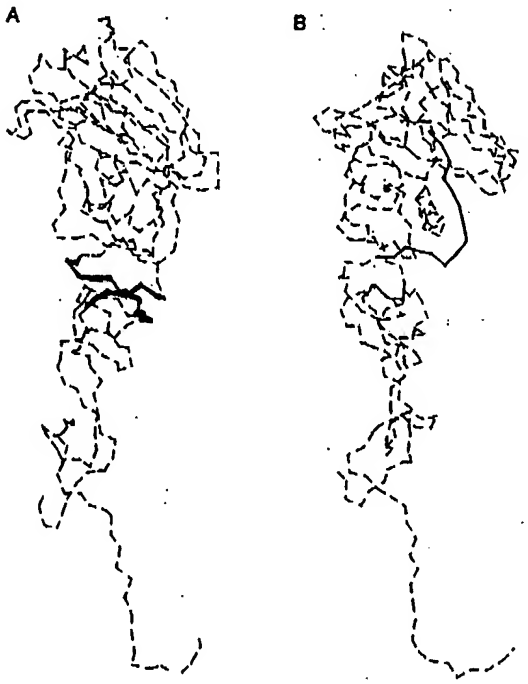

Fig. 1. Simplistic diagrams depicting (A) conformational and (B) sequential (or linear) determinants. The atomic coordinates of the influenza A/Hong Kong HA1 crystallographic structure (*13*) were rotated until prototypical regions recognized by eye were portrayed with Evans-Sutherland graphics by means of the GRAMPS and GRANNY programs of Donnell and Olson (*59*) and Connolly (*60*) to emphasize specific molecular features. The two panels show HA1 at different rotations around its vertical axis. The relevant regions are shown by the solid lines [in bold face in (A)]; the rest of the molecule is shown by dashed lines along the α-carbon backbone. (A) The molecule shows a conformational determinant—two regions of the HA1 polypeptide chain (residues 50 to 55 and 270 to 280) are far apart in the linear protein sequence but intimately related with each other through a disulfide linkage. The chains from the two regions track each other for several residues. One might expect that an antibody response to this region would probably involve contacts with both peptide chains. This appears to be the case during natural infection; this is HA1 site C of (*14*). However, antiserums to synthetic peptides representing either chain are capable of precipitating HA1 and, in fact, neutralizing the influenza virus. (B) The molecule shows a sequential determinant. The conformation of the peptide chain (residues 255 to 268) in the region shown by the solid line is mostly determined by the particular residues of the chain itself because no other close intramolecular contacts are apparent (nor are there any contacts with HA2 or other HA1 + 2 protomers in the trimeric form of the hemagglutinin). A synthetic peptide might well be expected to antigenically simulate a region such as this. These two extremes are simplifications that help to clarify abstract discussions of protein structure and immunogenicity. In reality, lengthy sequential regions of a protein usually have somewhat extensive contacts with other regions of the protein. However, linear peptides can often mimic the intact protein structure by eliciting a protein-reactive antiserum; therefore, some of the forces that mold global protein structure are already contained in relatively short peptide regions. Clearly, antibody molecules react with a number of spatially related atoms in the antigen molecule, and, in that sense, all determinants are conformational. Recent observations indicate that many peptides from within a protein molecule can act as antigenic determinants; thus many determinants are also sequential. The conformational-sequential nomenclature is probably no longer a useful one.

following simple rules, one can in general select peptides that will elicit antibodies reactive with intact proteins.

In retrospect, it is easy to understand why, in terms of their chemistry, synthetic peptides have been so successful as immunogens in these and several other recent studies and why it was previously believed that they would be ineffective and were therefore not vigorously studied. All of the effective peptides from the influenza HA1 study (*6*) could be shown, when matched to the crystallographic structure of the protein, to represent regions exposed to the solvent. This may be due, in part, to the fact that they contain polar residues. Polar residues may also provide components of antigenic determinants capable of forming strong electrostatic interactions with antibody. In addition, proline residues may be important because they occur at bends in the peptide chain, often at "corners" exposed to the solution. Because proline residues have an imide rather than an amide bond, several atoms surrounding a proline residue have a fixed three-dimensional relation to one another, whether in a short peptide or in a complete protein structure, and, as such, will be recognized equally by antibody in either situation. Furthermore, residues on the two sides of a proline tend to be near each other because the peptide chain more or less turns back on itself at *cis* proline kinks. This produces a two-chain structure which might be thought of as a minor "conformational" determinant, albeit formed by residues not too far from one another in the linear chemical formula of the protein. Although these explanations for the effectiveness of peptides with certain chemical properties are somewhat speculative, it is now clear that, in solution, peptides and proteins are often similar. Peptides in solution probably attain conformations dictated by their chemical makeup which resemble those that occur in native protein structures. Furthermore, proteins are probably not static structures, always closely resembling their crystalline form; in fact, they may exhibit their various linear domains in many conformations. In support of these notions, Niman and Lerner (*20*) have shown that 50 percent of the monoclonal antibodies selected for reactivity with one of the peptides in the influenza HA1 study also react with HA1 protein.

In contrast, antiserums raised against native HA1 do not react with any of the 20 HA1 peptides (*6*). Since these peptides span 75 percent of the HA1 primary sequence, including all of the known, mapped determinants, it is clear that most, if not all, of the immune response against native HA1 is directed against determinants not mimicked by short linear peptides. This observation is consistent with the body of data showing that antigenic determinants on intact proteins are largely conformational (*3*). Earlier investigators reasoned that because the determinants were conformational, only rarely would a linear peptide be an effective antigen; therefore synthetic peptides would not have general utility. The influenza HA1 study shows that pieces of a protein can elicit antibodies reactive with the whole protein which the whole protein itself cannot elicit (*6*). Moreover, the studies on HB_1Ag, influenza HA1, MuLV polymerase, and rabies glycoprotein demonstrate that these peptides are not difficult to select (*6*, *17–19*). On the basis of the rules formulated in the HB_1Ag study, one can select one or two peptides from a protein sequence with relatively high confidence of being able to elicit a protein-reactive antiserum, an important economic consideration both in time and research dollars.

Protein-Reactive Peptide Antibodies as Reagents for Molecular Studies

Synthetic peptides that elicit reagents capable of reacting with proteins of known primary sequence can be used to establish identity between a protein sequence and the protein itself. Many sequences are of a hypothetical nature, having been deduced from a gene sequence, and an experiment with a synthetic peptide may be more powerful for demonstrating colinearity between a protein and nucleic acid than a series of successful genetic experiments. Antiserums to peptides have been used to detect the putative products of the MuLV polymerase (*18*) and envelope (*7*) genes as well as the transforming genes of the Moloney sarcoma (*21*), simian sarcoma (*22*), feline sarcoma (*23*), avian myeloblastosis (*24*), SV40 (*7*), and polyoma (*25*) viruses. Such antiserums have also been used in similar studies of six human adenovirus 2 (*26*) transforming genes, the mouse mammary tumor virus long terminal repeat (LTR) (*27*), and several messenger RNA's from rat brain (*28*). Peptides from the NH_2- and COOH-terminals as well as from the middle of the adenovirus transforming proteins elicit protein-reactive serums. Antibodies to peptides have been used to track the processing of polyprotein precursors in cells infected with poliovirus (*29*), influenza (Fig. 2), and MuLV (*8*, *18*). Because of the predetermined specificity of the antiserums elicited by synthetic peptides, the proteins reacting with them are known to carry the specific peptide sequence. These reagents are particularly powerful when used as sets for tracking simultaneously the fate of various regions of a protein precursor (as shown for influenza HA in Fig. 2 and discussed below for the leukemia virus *pol* product). They are also useful in identifying alternative exon usage [as has been shown with immunoglobulin class D genes (*30*) and the adenovirus 2 E1A transcription unit (*31*)]. With the adenovirus 2 E1B transcription unit synthetic peptides have been used to show that the 53K and 19K protein products are translated in different triplet reading frames (*26*).

When the existence of a protein has been demonstrated, one wants to know its cellular location. Antiserums to a peptide from the COOH-terminus of the vesicular stomatitis virus (VSV) G protein do not react with intact infected cells, although polyclonal antiserums to the G protein do. When the cells are opened by treatment with detergent, a strong reaction at the inner surface of the membrane is detected with the antiserum to the COOH-terminal peptide (*32*). These experiments indicate that the VSV G protein spans the plasma membrane with its COOH-terminus protruding into the cytoplasm. In immunofluorescence studies of fixed, transformed fibroblast cells, antiserums to a COOH-terminal peptide from the Rous sarcoma virus (RSV) transforming protein were used (*33*). Fluorescence was codistributed with vinculin at cell-cell contact sites. Antiserum to an interior RSV *src* peptide [which cross-reacts with the endogenous cellular *src* as well as the transforming proteins of Fuginami (*fps*) and Y73 (*yes*) viruses] reacted with the focal adhesion plaques of RSV-transformed rat cells (*34*). Therefore, these reagents, derived by using the sequence of a protein (or gene), are capable of indicating in detail the cellular location of a protein.

Another use for peptide antiserums is the correlation of structure with function. If an antiserum to a peptide perturbs an assayable protein function, a protein containing the peptide sequence is implicated. Antiserum to a peptide present in the middle T sequence of polyoma virus inhibits protein kinase function in vitro (*35*), and an antiserum to a peptide present in the feline sarcoma virus *fes* gene sequence also inhibits protein kinase activity (*23*). Therefore, each of these proteins must have protein kinase activity: the predetermined specificity of the reagent allows one to rule

out the often cited caveats about proteins and enzymatic activities that happen to coincide. Antiserums against several peptides in the NH_2-terminal half of the putative MuLV *pol* polyprotein inhibit reverse transcriptase activity, whereas some peptides in the COOH-terminal portion inhibit a virion-associated endonuclease activity (*18*). The two sets of inhibiting antiserums immunoprecipitate two different proteins in infected cells, as well as their common precursor. Therefore, not only do the peptide antiserums establish the colinearity between the *pol* protein precursor and its nucleic acid sequence, but they identify the products of protein maturation and assign them an enzymatic activity. Antiserum to a COOH-terminal peptide predicted from the poliovirus replicase gene sequence precipitates the core protein p63 and its precursors and inhibits replicase and polyuridylic acid polymerase activities in vitro, indicating that these two activities reside in p63 (*36*). One can thus imagine a range of experiments in which peptide antiserums will be used to modify specific behaviors of proteins because they bind to a specific predetermined region of that protein.

When neither function nor structure is known, antiserums to peptides can provide a means for purifying the bona fide protein. Walter and his colleagues (*37*) have coupled to Sepharose the purified immunoglobulin fraction of a peptide antiserum and used this to purify the middle T antigen of polyoma virus. The protein was eluted from the immunoaffinity column by competition with excess peptide. A two-cycle purification scheme in which one uses the immunoglobulin fractions of antiserums against two different peptides from a protein sequence might lead to a quite pure protein preparation (*37*). Immunoprecipitation and gel electrophoresis have been used to isolate a precursor of the MuLV envelope membrane anchor to a purity such that its NH_2-terminal sequence could be determined by radiochemical sequencing methods (*8*). It might eventually be possible to use the complement fixation properties of antigen-antibody complexes as an assay for protein purification after various chromatographic steps (*38*).

Although most work with synthetic peptides has been done with antiserums raised in laboratory animals, some investigators have used clonal populations of peptide-specific lymphocytes for their experiments. The studies of Niman and Lerner (*20*) on influenza HA1, and those of Gentry *et al.* (*34*) on the RSV transforming protein, have shown that a high proportion of hybridomas obtained by fusion of spleen cells from peptide-immunized donors secrete antibody reactive with the native protein. Such doubly specific reagents can be prepared in large quantities and are particularly useful for studies of fixed cells or tissues or for large-scale diagnostic work. Monoclonal populations of B-cell precursors, which can bind peptide but are not yet capable of antibody secretion, have been used to probe the immunological repertoire to determine whether antigen tolerance to host proteins is inherited or acquired (*39*). In the influenza system, Lamb and his co-workers (*40*) found that hemagglutin-specific T-cell populations from human donors responded to 12 of 12 HA1 peptides analyzed. One peptide, corresponding to the COOH-terminal 24 amino acids, appeared to be immunodominant in the sense that three of four T-cell clones were specific for determinants within this sequence.

On the technical side, antibody titer is a relevant concern. Some peptides seem highly immunogenic, sometimes even without being coupled to carrier protein, eliciting antiserums of high titer after a 35-day immunization procedure in which three doses, each of 200 micrograms of peptide, are administered (*5*, *6*). Other peptides seem to require several months and several booster injections to achieve a reasonable titer. A second immunization procedure in which the peptide is coupled to a carrier protein different from that used in the initial series of inoculations seems to be very effective for eliciting a peptide-specific rather than a carrier-specific response (*28*). For many, but not all, peptides we have examined, glutaraldehyde coupling seems to be more effective than coupling through cysteine residues. However, glutaraldehyde modifies lysine NH_2 groups and hence can significantly interfere with key amino acids in some peptides. To increase the activity of antiserums, and to reduce background activities, purified immunoglobulin can be specifically enriched by adsorption to the antigenic peptide coupled to a solid support (*25*). The immunoglobulin is then released by a chaotropic agent or removed from the column by competition with excess peptide, then dialyzed in the presence of a chaotropic agent.

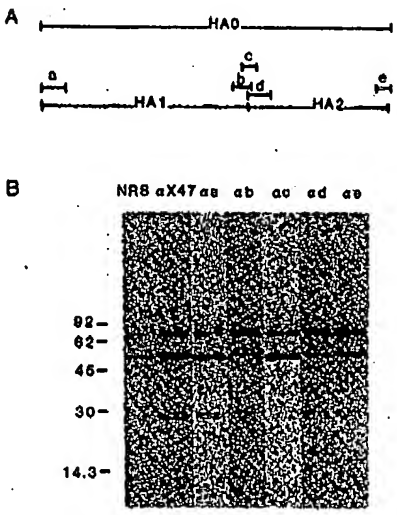

Fig. 2. Illustration of the power of sets of peptide antiserums used to track distinct regions of a protein simultaneously. (A) The influenza hemagglutinin precursor (HA0) and eventual products (HA1 and HA2). The synthetic peptides a to e, located within the HA0 sequence as shown (NH_2- and COOH-terminal, flanking and spanning the HA1-HA2 junction), were coupled to a carrier and used to immunize rabbits as described (*5*–*7*). (B) The resulting antiserums (αa, αb, αc, αd, αe) and a positive control serum (antibody to X47 influenza virus, αX47) and a negative control serum (normal rabbit serum, *NRS*) were used to precipitate extracts of [^{35}S]methionine-labeled influenza X47 virus-infected cells. The five antiserums to the peptides and αX47 precipitated the HA0 molecule (this precursor is not appreciably cleaved during infection of cell lines), whereas the other precipitated proteins were nonspecific (sticky) since they also appeared in the normal control lane.

Synthetic Peptides as Ideal Vaccines

For our studies we have used the immunological targets of infectious pathogens with the goal of eventually applying our findings to the protection of humans and other animals from disease by vaccination with synthetic peptides. Indeed, previous workers have shown that peptides of natural or synthetic origin can elicit antiserums capable of neutralizing or binding to virus or bacteriophage in vitro [for a review, see (*41*)]. The coat proteins of tobacco mosaic virus (TMV) (*42*) and more recently bacteriophage MS2 (*43*) were fragmented with proteases, and fragments that reacted with neutralizing serums to whole virus were identified and isolated biochemically or chemically synthesized. Antiserum to the natural COOH-terminal six residues of TMV had the ability to bind to the virus in vitro and abrogate its infectivity. Antiserums to a synthetic MS2 peptide reduced this bacteriophage's infectivity when a secondary antibody to the antibody or "sandwich" reaction was used (no primary neutralization was demonstrated). Similarly, Audibert *et al.* (*44*) demonstrated that antiserums against a synthetic peptide from the diphtheria toxin was capable of

inhibiting toxin activity in vitro. Our studies on HB_sAg (*17*) and influenza HA1 (*6*) showed that it is possible to select, on the basis of simple chemical properties of regions within viral proteins, peptide immunogens that induce virus-specific antibody. However, these studies did not establish the relation between the capacity to bind virus and the capacity to neutralize infection. What was needed was a demonstration of active protection from a virulent disease in an immunized animal.

In terms of neutralization, there seem to be two classes of virus—those that can be neutralized by antiserum to related viruses and those that are quite subtype-specific in their cross-reactivity. Protective peptides from the first class might cluster in regions whose sequences are conserved across related viruses, whereas protective peptides from the second class may reside in the variable regions. Although it is not necessarily the case that the serum of peptide-protected animals will mimic that of animals protected by classical vaccines, we have proceeded with the thought that reproducing the known serology may be important. The most extensive studies to date have been with four viruses: foot-and-mouth, influenza A, hepatitis B, and rabies.

Seven peptides were chosen (*45*) from the translated nucleic acid sequence of the type O foot-and-mouth disease virus (FMDV) VP1 protein (*46*). The VP1 was thought to be a target for neutralizing antibodies, because treatment of this picornavirus with trypsin cleaves only VP1 and this is sufficient to cause the virus particle to lose its immunizing activity (*47*). Rabbits inoculated with each of the seven peptides made strong antipeptide responses, but antiserums to only three of the peptides were capable of neutralizing FMDV in vitro (*45*). The neutralization was specific to type O virus, as expected from serological considerations. In addition, guinea pigs immunized with a single dose of either of two of the effective peptides (in complete Freund's adjuvant or alum) and challenged with live virus were protected. Thus, a synthetic peptide can elicit a protective immune response. Indeed, one of the peptides that was 20 residues long was as good at protecting the animals from later challenge with 10^4 infectious doses (ID_{50}) of FMDV as inactivated virus particles, the classical vaccine (*45*). Antiserums raised in rabbits or cattle against peptides from the equivalent region of other FMDV strains neutralize their appropriate strain of virus. These results are particularly encouraging because VP1 itself, whether purified from virus or made by genetically engineered bacteria, is a poor immunogen which thus far has not lent itself to the production of a subunit vaccine.

The antipeptide serums to influenza HA1 peptides (*6*) have been assayed for virus neutralization. Antibodies to 6 of the 20 peptides described in our original structural study neutralized the cytopathic effect of influenza virus on MDCK (canine kidney) cells (*48*). Subsequently, six more peptides were synthesized filling in the regions missed in the original study, including the HA1-HA2 junction and the other chain of the site C of Wiley *et al.* (*14*). Antiserums to five of these six peptides neutralize the cytopathic effect in vitro (*48*). In addition, immunizations with several combinations of peptides from the HA1 sequence protect mice from death caused by mouse-adapted influenza virus. However, only two of the five peptides thus far injected individually confer active immunity. Müller *et al.* (*49*) report neutralization of the cytopathic effect with antiserum to one influenza peptide and inhibition of virus growth in vivo by immunization with that peptide. Although it is too soon to make generalizations about structure, some statements about broad-range neutralization as measured in vitro by Alexander *et al.* (*48*) are possible. Sequences in the site C region (defined by a disulfide bond between residues 52 and 277 in the native HA1) (*14*), the COOH-terminal region of HA1 and the HA1-HA2 junction are relatively conserved in the several type A influenza hemagglutinins that have been sequenced. Antiserums to peptides from each of these regions neutralize not only viruses of the H_3 subtype (from which the sequence was derived) but also H_1 and B viruses (H_2 has not been tested). Normally, antiserums against one subtype of influenza virus do not protect against other subtypes. Thus with peptides it may be possible to construct a vaccine with a broad range of specificity not attainable with the intact protein.

The only relevant model for human hepatitis B is the chimpanzee, so direct studies of protection have not been done; however, extensive serological studies have been conducted in vitro. Prince *et al.* (*50*) demonstrated that a synthetic peptide corresponding to HB_sAg residues 138 to 149 inhibited the binding of hybridomas against the a and d but not the y subdeterminants of HB_sAg. In apparently conflicting studies carried out in the chimpanzee, Gerin *et al.* (*51*) report that a peptide spanning the region 110 to 137 contains the a and y but not the d specificity (d and y are thought to be allelic). This result is consistent with the results of protein chemistry studies of Peterson *et al.* (*52*), which show the y/d variation to occur at residues 131 and 134. This is supported by our studies showing that a synthetic peptide corresponding to residues 110 to 137 of the d version elicits d-specific antiserum. Bhatnagar *et al.* (*53*) have recently reported that a synthetic peptide covering the region 139 to 147 carries the a but not the y/d determinant. The reason for the discrepancy between the results of Prince *et al.* (*50*) and all of the others is not known, but whichever studies are correct, peptides seem to be able to duplicate serologically important antigenic determinants and elicit immune responses in primates corresponding to those seen during infection. No data on neutralization of the virus or active protection against infection are yet available.

In collaboration with colleagues at the Wistar Institute (*19*), we have investigated the efficacy of synthetic peptides in protecting animals from rabies virus infection. The purified rabies viral glycoprotein is capable of conferring protection from rabies virus, and all neutralizing serums and hybridomas precipitate this protein (*54*). Eighteen peptides, representing 56 percent of the primary rabies glycoprotein sequence as deduced from the nucleic acid sequence of its gene (*55*), were originally synthesized and antibodies were raised against these in rabbits and mice (*19*). All 18 peptides elicited antiserums that bound both rabies glycoprotein and virus as measured in ELISA assays. However, none was capable of neutralizing the virus in vitro or of protecting mice or dogs from challenge from live virus. More recently, it has been shown that the three cyanogen bromide fragments of the glycoprotein capable of eliciting protection were not well covered by the peptide selection (*56*), hence experiments directed at synthesizing these particular regions are under way. So, while it was easy to pick 18 out of 18 peptides that would elicit binding antibodies, none of these seems capable of protecting against the virus.

Clearly, the studies of FMDV and influenza virus show that one can protect an animal from infection with live virus by injecting synthetic peptides corresponding to parts of the proteins that are the normal targets of neutralizing antibodies. However, the studies with rabies and influenza virus demonstrate that binding antibodies, while easy to elicit, are not necessarily neutralizing antibodies. From the medical-veterinary point of view, one must either resort to synthe-

sizing a series of overlapping peptides covering a complete protein sequence (now a technically feasible, if not particularly elegant, approach) or perform supporting biological experiments to determine which regions of a protein contain its neutralization sites. Although sites capable of interacting with antibody are probably located on the entire surface of the viral protein with frequent representation in its primary linear amino acid sequence, and although such sites can probably be easily mimicked by synthetic peptides, sites susceptible to neutralization seem much less frequent and have not yet been characterized by a set of simple chemical properties. Nonetheless, once the neutralization sites are identified, synthetic peptides seem to be suitable substitutes for whole proteins or viruses (and presumably other infectious agents). For some sites, they may be more effective immunogens than whole protein because they can elicit specificities that whole proteins cannot.

If one projects this technology from prophylaxis to immunological therapy, synthetic peptides from proteins unique to tumor cells may be suitable immunogens for treating neoplasia. The peptides may be from proteins unique to the clonal population of transformed cells or, in tumors of tissues that are dispensible (such as prostate or thyroid), the peptides may be merely tissue-specific, cell-surface markers.

Conclusions

The desirability of synthetic vaccines of known potency and side effects has been recognized for many years. Some of the drawbacks of currently available inactivated virus vaccines are that they are not stable without refrigeration and sometimes contain incompletely inactivated viruses capable of starting minor epidemics (*57*). Vaccination with a substance free of any biological contamination introduced in its production (either by virus grown in cell culture or in genetically engineered proteins expressed by *Escherichia coli* protein synthesis factories) and unable to cause any virus-related pathology because of incomplete virus inactivation or imperfect attenuation is the logical goal of protection from infectious disease.

Certainly this technology has not been developed to the level of sophistication that is required for the widespread use of synthetic vaccines in humans and other animals. The adjuvants and carriers used in the studies described in this article are in general much too harsh for human use, although alum (used in the FMDV study) is suitable. The work of Chedid, Audibert, and Langbeheim and their colleagues (*58*) may indicate a possible direction for more suitable and ultimately totally defined vaccines. Suitable doses, the possibility of using modified peptides or combinations or polymers of peptides, and the various routes of injection need to be worked out. But a major theoretical obstacle has been overcome in that solving and then synthesizing complex conformational determinants no longer seems necessary.

Now that it has been shown that protection by synthetic peptides is possible, that such peptides can be at least as effective as biological vaccines, that new protein sequences are rapidly being generated as a result of nucleic acid studies, and that the synthetic approach is economically quite feasible (whether it requires the synthesis, by brute force, of all fragments of a protein or directed biochemical-immunological experiments), research to find the best peptides, the best adjuvants, and the best carriers is likely to become an important priority. As the research progresses, and as synthetic vaccines become commonplace, the somewhat tedious processes used today will become streamlined and consequently much more economical.

References and Notes

1. F. Sanger, *Science* 214, 1205 (1981); W. Gilbert, *ibid.*, p. 1305.
2. B. Yelverton, D. Leung, P. Weck, P. W. Gray, D. V. Goeddel, *Nucleic Acids Res.* 9, 731 (1981); D. G. Kleid, D. Yansura, B. Small, D. Dowbenko, D. M. Moore, M. J. Grubman, P. D. McKercher, D. O. Morgan, B. H. Robertson, H. L. Bachrach, *Science* 214, 1125 (1981); P. Charnay, M. Gervais, A. Louis, F. Galibert, P. Tiollais, *Nature (London)* 286, 893 (1980).
3. K. Landsteiner, *J. Exp. Med.* 75, 269 (1942); M. J. Crumpton, in *The Antigens*, M. Sela, Ed. (Academic Press, New York, 1974), p. 1; E. Benjamini, J. D. Young, M. Shimizu, C. Y. Leung, *Biochemistry* 3, 1115 (1964); M. Z. Atassi and B. J. Suplin, *ibid.* 7, 688 (1968); R. Arnon, R. Maron, M. Sela, C. B. Anfinson, *Proc. Natl. Acad. Sci. U.S.A.* 68, 1450 (1971); J. J. Cebra, *J. Immunol.* 86, 205 (1961).
4. R. A. Lerner, *Nature (London)* 299, 592 (1982).
5. In practice, peptides are synthesized according to the high repetitive yield methods developed by Merrifield and his colleagues [A. Marglin and R. B. Merrifield, *Annu. Rev. Biochem.* 39, 841 (1970)] by generally available automatic synthesizers or by hand. They are usually used without further purification, as possible contaminants are either peptides with trivial sequence variations of that designed or they are multimeric forms. The peptides are coupled to a carrier protein (keyhole limpet hemocyanin, edestin, thyroglobulin, and bovine serum albumin have been used) and injected into animals (usually in complete Freund's adjuvant). Some peptides are immunogenic without being coupled to a carrier. Serum from immunized animals is assayed for its ability to bind peptide antigen, either in solution or in a solid phase [enzyme-linked immunosorbent (ELISA)] assay. Positive sera are then tested against native proteins [see (*6*)].
6. N. Green, H. Alexander, A. Olson, S. Alexander, T. M. Shinnick, J. G. Sutcliffe, R. A. Lerner, *Cell* 28, 477 (1982).
7. J. G. Sutcliffe, T. M. Shinnick, N. Green, F.-T. Liu, H. L. Niman, R. A. Lerner, *Nature (London)* 287, 801 (1980); O. Walter, K. H. Scheidtmann, A. Carbone, A. P. Laudano, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* 77, 5197 (1980).
8. N. Green, T. M. Shinnick, O. Witte, A. Ponticelli, J. G. Sutcliffe, R. A. Lerner, *Proc. Natl. Acad. Sci. U.S.A.* 78, 6023 (1981).
9. R. A. Lerner, J. G. Sutcliffe, T. M. Shinnick, *Cell* 23, 309 (1981).
10. There is nothing intrinsically different between protein sequences derived by translating nucleic acid sequences according to the rules of the genetic code and those empirically determined (except that an empirically determined protein sequence has basis in experiment), but most new protein sequences are determined by nucleic acid sequencing.
11. M. Pasek, T. Goto, W. Gilbert, B. Zink, H. Schaller, P. McKay, G. Leadbetter, K. Murray, *Nature (London)* 282, 575 (1979); W. Minjou, M. Verhoeyen, R. Devos, E. Saman, R. Fang, D. Huylebroeck, W. Fiers, G. Threlfall, C. Barber, N. Carey, S. Emtage, *Cell* 19, 683 (1980).
12. G. N. Vyas, K. R. Rao, A. B. Ibrahim, *Science* 178, 1300 (1972).
13. I. A. Wilson, J. J. Skehel, D. C. Wiley, *Nature (London)* 289, 366 (1981).
14. D. C. Wiley, I. A. Wilson, J. J. Skehel, *ibid.*, p. 373.
15. W. G. Laver, G. M. Air, T. A. Dopheide, C. W. Ward, *ibid.* 283, 459 (1980); R. G. Webster and W. G. Laver, *Virology* 104, 139 (1980).
16. D. L. Peterson, I. M. Roberts, G. N. Vyas, *Proc. Natl. Acad. Sci. U.S.A.* 74, 1530 (1977).
17. R. A. Lerner, N. Green, H. Alexander, F.-T. Liu, J. G. Sutcliffe, T. M. Shinnick, *ibid.* 78, 3403 (1981).
18. T. M. Shinnick and J. G. Sutcliffe, unpublished observations.
19. J. G. Sutcliffe, M. Kiel, T. J. Wictor, D. Lopes, H. Koprowski, unpublished observations.
20. H. L. Niman and R. A. Lerner, unpublished observations.
21. J. Pankoff, I. M. Verma, T. Hunter, *Cell* 29, 417 (1982).
22. K. C. Robbins, S. G. Devare, B. P. Reddy, S. A. Aaronson, *Science* 218, 1131 (1982).
23. S. Sen, R. Houghten, C. Scherr, A. Sen, personal communication.
24. J. Lukacs and M. Baluda, personal communication.
25. G. Walter, M. A. Hutchinson, T. Hunter, W. Eckhart, *Proc. Natl. Acad. Sci. U.S.A.* 78, 4882 (1981).
26. M. Green, K. Brackmann, J. Symington, personal communication; P. B. Branton, S. P. Yee, D. T. Rowe, personal communication.
27. N. Fasel, D. Owen, E. Buetti, H. Diggelmann, personal communication.
28. J. G. Sutcliffe, R. J. Milner, T. M. Shinnick, R. A. Houghten, F. E. Bloom, unpublished observation.
29. M. H. Baron and D. Baltimore, *Cell* 28, 395 (1982); B. L. Semler, C. W. Anderson, R. Hanecak, L. F. Dorner, E. Wimmer, *ibid.*, p. 405.
30. T. M. Shinnick and F. Blattner, unpublished observation.
31. L. Feldman and J. Nevins, personal communication.
32. M. C. Willingham, R. Schlegel, I. Pastan, personal communication.
33. E. A. Nigg, B. M. Sefton, T. Hunter, G. Walter, S. J. Singer, *Proc. Natl. Acad. Sci. U.S.A.* 79, 5322 (1982).
34. L. E. Gentry, L. R. Rohrschneider, J. E. Casnelli, R. Beer, E. G. Krebs, personal communication.
35. B. Schaffhausen, T. L. Benjamin, L. Pike, J. Casnellie, E. Krebs, *J. Biol. Chem.*, in press.
36. M. H. Baron and D. Baltimore, *J. Virol.* 43, 969 (1982).
37. G. Walter, M. A. Hutchinson, T. Hunter, W. Eckhart, *Proc. Natl. Acad. Sci. U.S.A.* 79, 4025 (1982).
38. R. Tijan, *Cold Spring Harbor Symp. Quant. Biol.* 43, 655 (1979).
39. R. Jemmerson, P. Morrow, N. Klinman, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* 41, 420 (1982).
40. J. R. Lamb, D. D. Eckels, P. Lake, J. N. Woody, N. Green, *Nature (London)* 300, 66 (1982).
41. R. Arnon, *Annu. Rev. Microbiol.* 34, 593 (1980).
42. R. A. Anderer, *Biochim. Biophys. Acta* 71, 246 (1963).
43. H. Langbeheim, R. Arnon, M. Sela, *Proc. Natl. Acad. Sci. U.S.A.* 73, 4336 (1976).
44. R. Audibert, M. Jolivet, L. Chedid, J. E. Alout, P. Boguet, P. Rivielle, O. Siffert, *Nature (London)* 289, 593 (1981).
45. J. L. Bittle, R. A. Houghten, H. Alexander, T. M. Shinnick, J. G. Sutcliffe, R. A. Lerner, D. J. Rowlands, F. Brown, *Nature (London)* 298, 30 (1982).
46. C. Kurz, S. Forss, H. Kupper, K. Strohmaier,

H. Schaller, *Nucleic Acids Res.* 9, 1919 (1981).
47. K. Strohmaier, R. Franze, K.-H. Adam, *J. Gen. Virol.* 59, 295 (1982).
48. S. Alexander, H. Alexander, N. Green, R. A. Lerner, personal communication.
49. G. Müller, M. Shapira, R. Arnon, *Proc. Natl. Acad. Sci. U.S.A.* 79, 569 (1982).
50. A. M. Prince, H. Ikram, T. P. Hopp, *ibid.*, p. 579.
51. J. L. Gerin *et al.*, *ibid.*, in press.
52. D. L. Peterson, N. Natu, F. Gavilanes, *J. Biol. Chem.* 257, 10414 (1982).
53. P. K. Bhatnagar, E. Papas, H. E. Blum, D. R. Milich, D. Nitecki, M. J. Kareis, G. N. Vyas, *Proc. Natl. Acad. Sci. U.S.A.* 79, 4400 (1982).
54. J. H. Cox, B. Deitzschold, L. O. Schneider, *Infect. Immun.* 16, 743 (1977).
55. A. Anilionis, W. H. Wunner, P. J. Curtis, *Nature (London)* 294, 275 (1981).
56. B. Dietzschold, T. J. Wiktor, R. MacFarlan, A. Varrichio, *J. Virol.*, in press.
57. J. Beale, *Nature (London)* 298, 14 (1982); A. M. Q. King, B. O. Underwood, D. McCahon, J. W. I. Newman, F. Brown, *ibid.* 293, 479 (1981).
58. L. Chedid, F. Audibert, A. Johnson, *Prog. Allergy* 25, 63 (1978); H. Langbeheim, R. Arnon, M. Sela, *Immunology* 35, 573 (1978); F. Audibert, M. Jolivet, L. Chedid, R. Arnon, M. Sela, *Proc. Natl. Acad. Sci. U.S.A.* 79, 5042 (1982).
59. T. J. O'Donnell and A. J. Olson, *Computer Graphics* 15, 133 (1981).
60. M. Connolly, unpublished results.
61. We acknowledge the efforts of our co-workers who contributed to designing the experiments and collecting the data to which we refer herein, and thank many of our colleagues for sharing their data in advance of publication. We also thank A. Olson for producing Fig. 1 and R. Ogata for comments on the manuscript. Portions of this work were supported by grants from the American Cancer Society (NP-359) and the National Institutes of Health (R01 AI 18509). This is paper No. 2824 of the Research Institute of Scripps Clinic.

Protein Engineering

Kevin M. Ulmer

Summary. The prospects for protein engineering, including the roles of x-ray crystallography, chemical synthesis of DNA, and computer modeling of protein structure and folding, are discussed. It is now possible to attempt to modify many different properties of proteins by combining information on crystal structure and protein chemistry with artificial gene synthesis. Such techniques offer the potential for altering protein structure and function in ways not possible by any other method.

In the last decade, genetic engineering technology has been developed to the point where we can now clone the gene for essentially any protein found in nature. By precise manipulation of the appropriate regulatory signals we can then produce significant quantities of that protein in bacteria. Recent advances in chemical synthesis of DNA now permit virtually unlimited genetic modification, and offer the prospect for developing protein engineering technology to create novel proteins not found in nature. By starting with the known crystal structure for a protein we would like to directly modify the gene to alter that structure in a predictable fashion, targeted to improve some functional property. At each stage we could verify the structural and functional changes that actually occurred and thereby refine and extend our predictive capability. Step by step, as we gain facility with this technique and learn the detailed rules that relate structure and function, we should be able to create proteins with novel properties which could not be achieved as effectively by any other method.

Rationale

Despite the fact that biochemists have characterized several thousand enzymes, there are only a handful that could be considered enzymes of commerce. Indeed, only a dozen enzymes have worldwide sales in excess of $10 million per year, and together they account for more than 90 percent of the total enzyme market (*1*). Frequently the limiting factor in the industrial use of an enzyme has simply been the high cost of isolating and purifying adequate amounts of the protein. Part of the solution to this problem lies with the ability of genetic engineers to greatly amplify the production of specific enzymes in microorganisms, but beyond cost there are often other limitations to the broader use of enzymes which stem from the fact that the desired industrial application is far removed from the physiological role normally played by the enzyme. In particular, industrial applications require generally robust enzymes with a long half-life under process conditions. Frequently the desired substrate or product is somewhat different from the physiological one, and often the chemical conditions for the reaction are decidedly nonphysiological, ranging to extremes of *p*H, temperature, and concentration. If enzymes are to be more widely used as industrial catalysts, we must develop methods to tailor their properties to the process of interest. The list of properties of enzymes we would like to be able to control in a predictable fashion would include the following:

1) Kinetic properties including the turnover number of the enzyme and the Michaelis constant, K_m, for a particular substrate.

2) Thermostability and temperature optimum.

3) Stability and activity in nonaqueous solvents.

4) Substrate and reaction specificity.

5) Cofactor requirements.

6) *p*H optimum.

7) Protease resistance.

8) Allosteric regulation.

9) Molecular weight and subunit structure.

The solutions to these problems have included extensive searches for the best suited naturally occurring enzyme, mutation and selection programs to enhance the native enzyme's properties, and chemical modification and immobilization to obtain a stable and functional biocatalyst. From such work we know that all of these properties can in general be improved. Specific examples of what has been achieved by these methods and how protein engineering can build on this knowledge to yield still further improvements are cited below.

It is not uncommon to observe wide variations in properties such as turnover number, K_m, molecular weight, temperature optimum, thermostability, *p*H optimum, and *p*H stability among enzymes of the same type isolated from different

Kevin M. Ulmer is director of exploratory research at Genex Corporation, Science and Technology Center, 16020 Industrial Drive, Gaithersburg, Maryland 20877.